MBROLA Delexicaliser v. 1.0
A Tool for Creating Delexicalised Stimuli for Speech Perception Experiments
Technical Report and User's Guide

George Christodoulides

Language Sciences and Metrology Unit, Université de Mons,
Place du Parc 18, B-7000 Mons, Belgium
george@mycontent.gr

Abstract. This technical report presents the MBROLA Delexicaliser, a software tool for creating delexicalised stimuli (i.e. stimuli in which it is impossible for the listener to understand the lexical content) on the basis of existing recordings and time-aligned phonetic transcriptions. While the generated stimuli retain the prosodic characteristics (duration, pitch targets) of the initial recordings, each phoneme of the original stimulus is replaced with another phoneme from the same class. The tool generates synthesis files for the MBROLA multilingual speech synthesiser, which is free for non-commercial use. The resulting stimuli sound like speech in a pseudo-language similar to the original language and have similar prosody, but the lexical content is completely hidden. This technique is often more effective than low-pass filtering for de-lexicalisation.

Keywords: speech · phonetics · delexicalisation · experimental linguistics · software tool.

1 Introduction

In experimental linguistics and research on speech prosody, it is sometimes useful to construct synthesised speech samples (stimuli) that are delexicalised, i.e. samples where the lexical content is masked and impossible to recognise, and have a specific prosodic profile (duration of segments, pauses, intonation contours etc). In this technical report, we present a software tool that creates delexicalised versions of utterance recordings, based on a time-aligned phonetic transcription. The system, called MBROLA delexicaliser, replaces all phonemes in the transcription with others from the same class of phonemes (respecting the phonotactics rules supplied by the user), copies the segment durations and pitch contours of the original recordings and produces the stimuli by feeding this information to the MBROLA speech synthesiser.

The MBROLA speech synthesiser [3] is the result of a project initiated by the Faculté Polytechnique de Mons (Belgium), to obtain a set of speech synthesizers for as many voices, languages and dialects as possible, free of use for non-commercial and non-military applications. The MBROLA synthesiser is based on the concatenation of

diphones and allows for precise control of segment durations and pitch targets in the output. It is freely available for many computer operating systems.

The technical report is structured as follows. In Section 2 we will present the Language Definition File, by which the user defines the phonemes of a language, their classes and the constraints in their combinations (phonotactics) that the tool should follow. In Section 3 we will show how to prepare a recording and the associated transcription for use with the tool and in Section 4 we will present the use of the Delexicaliser tool to produce the input files for the MBROLA speech synthesiser. Section 5 describes the example files provided online, Section 6 outlines the perspectives for future development of the tool, and Section 7 contains license information.

2 Language Definition File

A language definition file is needed to operate the MBROLA Delexicaliser. A sample definition file is provided for French. The language definition file is a JSON document that contains the following elements: (a) the name of the language; (b) the name of the phoneme set used for the phonetic transcription (e.g. SAMPA); (c) a list of phoneme classes and their members (e.g. fricatives, plosives etc); and (d) a list of all possible diphones (occurrences of two contiguous phonemes that are allowed according to the language's phonotactics).

The Delexicaliser will substitute each phone in the original stimulus with another phoneme from the same phoneme group (e.g. a fricative with another fricative), always ensuring that the resulting diphones, to the left and to the right, are possible diphones based on the language description.

The sample file for French defines the following phoneme classes:

Phoneme Class Name	Phonemes (SAMPA)
plosives	p, b, t, d, k, g
fricatives	f, v, s, z, S, Z
nasals	m, n
liquids	l, R
glides	w, H, j
vowels	i, e, E, a, A, o, O, y, 2, 9
nasal vowels	e , a , o , 9
schwa	@

Table 1: Phoneme classes in the language definition file for French (example).

You can use this sample file as a template to produce a language definition file for your own language. You can also change the grouping of the phonemes into phoneme classes, depending on your research question. Grouping more phonemes into larger classes will generally lead to de-lexicalised stimuli that sound more different than the original stimuli.

3 Preparing the Original Files

The process outlined in this section is necessary only if you use the MBROLA Delexicaliser stand-alone GUI application. If you use the MBROLA Delexicaliser as a plug-in for Praaline [2], the plug-in performs these steps automatically.

You will need the recording file in wave format (.wav) and a corresponding Praat [1] TextGrid containing a phonetic transcription. Figure 1 shows a TextGrid with such a transcription on the first tier (the other tiers are not necessary for operating the Delexicaliser). The symbols used for the phonetic transcription must match those described in the Language Definition File.

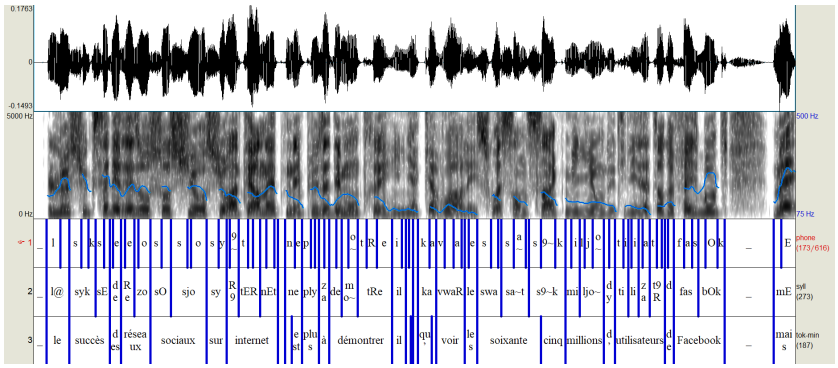

Fig. 1: A Praat TextGrid and the corresponding sound file. Note that the Delexicaliser only needs to have a "phones" tier.

You will also need to create a Pitch Tier file for your recording. To do so, follow these steps: open the recording in Praat and select the command "To Pitch (ac)..." from the menu "Analyse Periodicity" on the right-hand side of the Objects window. You can use the default settings or tweak the parameters for pitch detection (especially the lower and upper threshold). Praat will create a new Pitch object in the Objects window; select it and then select "Down to PitchTier" from the "Convert" menu on the right. A new PitchTier object will be created; select this Pitch Tier object and save it as a text file (menu "Save", command "Save as text file"). Repeat this process for all samples.

You should now have three files for each sample that you want to process: sample.wav (the recording), sample.TextGrid (the annotation file) and sample.PitchTier (the pitch tier).

You may wish to consider using a pitch stylisation tool, such as Piet Merten's Prosogram [5] in order to obtain a pitch tier that is smoother and corrected for common pitch detection errors such as octave jumps. This is the preferred option when using the MBROLA Delexicaliser as a plug-in for Praaline.

4 Generating the Delexicalised Stimuli

The next step is to load the files in the MBROLA Delexicaliser tool, which is shown in Figure 2. On the left, there is the list of files to be processed; you can add files using the (plus) toolbar button, remove files using the (minus) toolbar button, or add an entire folder using the (folder) toolbar button.

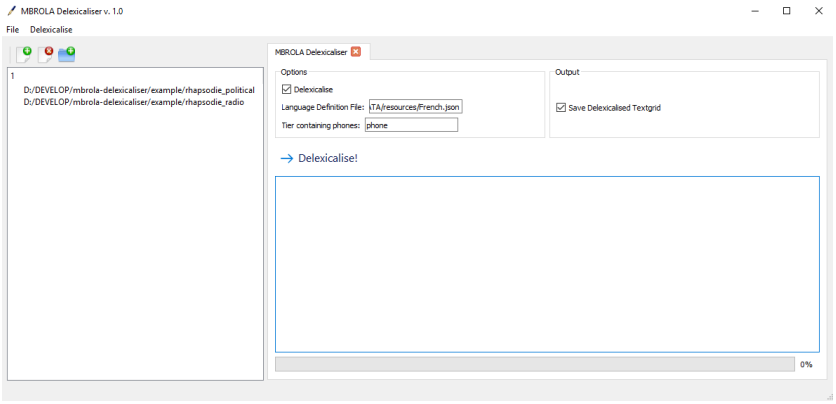

Fig. 2: The graphical user interface of MBROLA Delexicaliser under Windows.

You should provide the path to the language definition file (by default, it is on a folder called “resources” next to the Delexicaliser’s executable file) and the name of the tier that contains the phonetic transcription (by default it is “phone”). If you select the option “Delexicalise”, the tool will actually substitute phonemes; you can uncheck this option if you want to obtain a MBROLA synthesis (.pho) file containing the original segments, durations and pitch contours (e.g. to select the closest matching voice). For normal operation, you will want to select “Delexicalise”. Finally, you have the option to save a Praat TextGrid containing the modified phonetic transcription; all other tiers in the original TextGrids will be copied over.

When done, click on the button “Delexicalise”. For each sample file, the system will produce a file called sample_delexicalised.pho, that should be given as input to the MBROLA speech synthesis system. If the relevant option was selected, it will also produce a sample_delexicalised.TextGrid file.

You then need to download and set up the MBROLA speech synthesiser, as described on the project’s website, and generate the synthesised stimuli using the .pho files.

5 Example Files

We provide two example files for download. The samples were extracted from the Rhapsodie corpus of spoken French [4]. The first sample (rhapsodie_political) is a po-

litical speech (sample M2001 in the Rhapsodie corpus) and the second sample (rhapsodie_radio) comes from a radio broadcast (sample M2005 in the Rhapsodie corpus). The corresponding phonetic transcriptions and Pitch Tier files are already provided. The Delexicaliser tool has generated the rhapsodie_radio_delexicalised.pho and rhapsodie_political_delexicalised.pho files, which were used to synthesise the two corresponding wave files of the delexicalised stimuli.

6 Future Work

The Delexicaliser tool is under active development. We intend to make the following improvements:

- Add functionality to allow for delexicalisation by replacing syllables instead of individual phones, ensuring that phonotactics rules are respected and that syllables of similar characteristics (to the extent possible) are used. This will create smoother-sounding stimuli.
- Add functionality for pitch extraction (and simple smoothing); processing the files with Praat's pitch extraction algorithm will no longer be necessary.
- Copy over the intensity/loudness characteristics of the original recording to the delexicalised stimulus.

7 Download, License and Contact

You can download the MBROLA Delexicaliser at: https://github.com/praaline/mbrola-delexicaliser/releases

The MBROLA Delexicaliser tool is free, open-source software, distributed under the GNU General Public License (GPL) version 3.0. The source code is available on the GitHub repository:
https://github.com/praaline/mbrola-delexicaliser/

The MBROLA speech synthesiser (which is needed by this tool) is distributed under a license agreement that allows non-commercial, non-military use of the program and databases (voices). The license for MBROLA is available at:
https://tcts.fpms.ac.be/synthesis/

The MBROLA Delexicaliser is written in C++ and uses the LGPL version of the Qt Framework (www.qt.io).

You are welcome to contact the author if you find bugs, if you have suggestions, or if you want to contribute.

References

1. Boersma, P., Weenink, D.: Praat: doing phonetics by computer [computer programme]. Version 6.0.36 `http://www.praat.org` (2017)
2. Christodoulides, G.: Praaline: integrating tools for speech corpus research. In: LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings. pp. 31–34 (2014), `http://www.praaline.org`

3. Dutoit, T., Pagel, V., Pierret, N., Bataille, F., van der Vrecken, O.: The Mbrola project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In: Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96). vol. 3, pp. 1393–1396 (1996)
4. Lacheret, A., Kahane, S., Beliao, J., Dister, A., Gerdes, K., Goldman, J.P., Obin, N., Pietrandrea, P., Tchobanov, A.: Rhapsodie: a prosodic-syntactic treebank for spoken French. In: LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings (2014)
5. Mertens, P.: The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In: Proc. of Speech Prosody 2004, March 23–26, Nara, Japan. pp. 549–552 (2004)